

# Web Platform Providing Information and Simulation Game of Baseball

Chaewon Ko, Changwoo Shim, and Hyunchang Shin

Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do,  
Republic of Korea

**Abstract.** The BaseballMetrics is web platform for fans of baseball. The project is motivated by increasing interest of Sabermetrics and obstacles to enjoy baseball. Because of difficulty of sabermetrics, most fans feel hard about indicator about sabermetrics. Also, influx of new fans is decrease because of many hard terms about sabermetrics. And there are some obstacles to enjoy baseball. There are many breaktimes in baseball game. So, fans who watch baseball with online method feel boring. BaseballMetrics can be solution for this problem. In this platform, it will be information of sabermetrics in comprehensible words, and easy example. And there will be simulation game using sabermetrics. To make this game, we will design new algorithm and prediction model using sabermetrics. With this game, user can become familiar with sabermetrics. This service also serves many interesting functions for baseball fans. The platform will have the effect of making existing and new fans more interested in baseball.

**Keywords:** Baseball · Sabermetrics · Web Platform.

## 1 Introduction

By far, the most enthusiastic sports league among Koreans is Korean Baseball League, KBO. At the bottom of the ninth inning, a pitcher or batter entering a situation where one hit determines the game is enough to make us expect. Not necessarily in this situation, baseball makes us look forward to it, and it's fun to look forward to what will happen. So do coaches also rely on expectations for decisions during the game? The answer is maybe, no.

Baseball is called "The Sports of Record". This is because there are more records than other sports, and records alone can help you understand the content of the game and the player's ability. Why did baseball become a sport of record? Ironically, no sport is more uncertain than baseball. Whether there is a .200 hitter or a .300 hitter, the result is zero or one hit. Even if there is a 0% hitter, there is a chance of hitting a hit. In some cases, the batter gets on base because the catcher can't catch the ball even though he struck out, and in some cases, he runs too hard and is out at second base. In this uncertainty, the process of experts struggling to somehow exclude the element of luck and quantify expectations has

made baseball a sport of record. And this uncertainty makes us fall in love with baseball, whether we're fans, coaches or data analysts. However, the terms and numbers are so diverse that it is difficult for beginners to grasp them at a glance.

## 2 Motivation and Objective

### 2.1 Motivation

Sabermetrix actively introduces game theory and statistical methodology of social science into baseball to complement the poor parts of existing baseball records and to try a more academic and in-depth approach to baseball. There are many indicators that are unfamiliar to general baseball fans or difficult to intuitively interpret the figures.

Somewhere there are people who argue that there is no need for indicators to enjoy baseball. It's not wrong. Some people enjoy baseball with anticipation and disappointment or joy at every moment. There are also people who like the atmosphere of the baseball stadium and chicken and beer. Everyone has different ways and reasons to enjoy baseball. But obviously, there are many fans who are interested in baseball's records and want to understand the flow of the game and coaches' judgment.

As a sport of uncertainty, there is no correct answer to the coach's judgment. The director also chooses the most likely side, and sometimes makes judgments that depend on the senses. Baseball will be more enjoyable if you can understand and enjoy the basis of this judgment. You will be able to express your regret with an appropriate basis for the director's strange-looking judgment. Or you'll be able to understand that it might be.

Therefore, we want to provide baseball fans with the opportunity to get acquainted with Sabermetrics (the records and indicators that have been mentioned so far).

### 2.2 Objective

We propose a web platform to familiarize myself with Sabermetrics. We think rigid explanations are no different from Wikipedia and people will hate to read them. As a result, this platform came up with a platform that can be enjoyed in the form of mini-games, arouse natural curiosity, and provide answers. This mini-game will include indicators that were somewhat unfamiliar, and explanations.

Due to the nature of the web platform, it can provide several services and also provides a function to calculate the magic number. Magic number means the minimum multiplier required to finish the season in a specific ranking. Fans are curious about the magic number of the team they support, but sports articles

only cover the first and fifth place (to advance to the tournament) until the end of the season. In addition, it will increase the exposure of the game by inducing users to access the page.

### 3 Background and Relative Work

#### 3.1 Background

**What is Sabermetrics?** The project focuses on Sabermetrics, a concept rooted in the nature of baseball, where the significance of records is paramount. In baseball, various statistical records play a crucial role, serving as meaningful indicators of the abilities and performances of teams and players. Since the inception of baseball, records have been created, gaining widespread recognition as important factors in the sport.

These meaningful records are actively mass-produced and maintained across various professional leagues. Sabermetrics, as an empirical analysis of baseball statistics measuring in-game activity, serves as a tool for enthusiasts seeking objective insights. It emerged to provide a comprehensive evaluation of players in batting, pitching, and fielding. The emphasis is on expressing assessments in terms of runs or team wins, challenging the perception that older statistics are inherently less effective. Essentially, Sabermetrics offers a contemporary perspective for baseball enthusiasts to gain deeper insights into the game's intricacies.

**Purpose of Sabermetrics** Sabermetrics serves a range of purposes, commonly applied to assess both past and predict future player performance, crucial for decisions on awards and trades in baseball. Minor league experiences before reaching the majors introduce challenges in comparing player statistics. Sabermetricians address this by adjusting minor league stats, termed Minor-League Equivalency, allowing teams to evaluate a player's potential major league readiness based on performances in AA and AAA.

In the realm of applied statistics, sabermetrics is employed for comparing key performances, evaluating past player performances to determine market values, predicting future performances, and understanding a player's impact on team success or failure. This objective analysis aids decisions like signing or releasing players, considering their specific characteristics.

#### 3.2 Related Works

**Previous baseball measurement** The conventional gauge of batting prowess, hits divided by total at-bats, was criticized by Bill James and other sabermetrics pioneers for its limitations in overlooking alternative ways a batter reaches base. In contrast, on-base percentage (OBP) factors in walks and hit-by-pitches. Another critique of batting average is its failure to distinguish between types

of hits, leading to the introduction of slugging percentage (SLG), calculated by dividing the total bases of all hits by the total number of at-bats.

In pitching, the traditional measure, earned run average (ERA), calculates earned runs allowed per nine innings but fails to isolate the pitcher's individual skill from the performance of the fielders. Another classic metric, a pitcher's winning percentage, divides wins by total decisions (wins plus losses) but is heavily influenced by the team's overall performance, especially its scoring capabilities.

### **Advanced measurement**

- *wRC<sup>+</sup>*

*wRC<sup>+</sup>* modifies the Runs Created statistic by incorporating adjustments for significant external factors, such as ballpark variations or era influences. The adjustment ensures that a *wRC<sup>+</sup>* score of 100 represents league average, with a score of 150 indicating performance 50 percent above the league average. This metric quantifies run creation and standardizes it, allowing for comparisons among players in diverse ballparks and across different eras.

- WHIP

WHIP (Walks plus Hits per Inning Pitched) is a widely used statistic for assessing a pitcher's performance, indicating how effectively they've kept runners off the basepaths. It is calculated by summing a pitcher's walks and hits, then dividing by total innings pitched. Lower WHIP values generally denote superior pitchers, as they excel in preventing baserunners. However, WHIP doesn't distinguish how a hitter reached base, omitting consideration for the severity of hits like home runs. Certain factors, such as hit batsmen, errors, and fielder's choice, don't contribute to a pitcher's WHIP.

- BABIP

BABIP (Batting Average on Balls in Play) assesses a player's batting average specifically on balls hit into the field, excluding home runs and strikeouts. It provides context when evaluating both pitchers and hitters, with a league average around .300. Pitchers with high BABIP may regress to the mean, experiencing better results over time, and vice versa. Skill can influence BABIP, as some pitchers induce weak contact, and certain hitters excel at producing hard-hit balls. For instance, Clayton Kershaw consistently maintains a low BABIP, while Mike Trout has a higher career BABIP due to his hitting prowess.

- VORP

Value over replacement player (VORP) is a statistic that highlights a hitter or pitcher's contribution to their team relative to a replacement-level player—a player considered average in fielding and below average in hitting. A replacement player operates at "replacement level," representing the expected performance of an average team when seeking a player replacement

at minimal cost, often referred to as "freely available talent." The value of VORP lies in its ability to gauge the marginal utility of individual players. In contrast to other statistics that compare players to the league average, which is beneficial for analyzing performance across different eras. In the formula, 'IP' and 'RA9' respectively refers to 'innings pitched' and 'run average'.

## Existing services

- PECOTA

The acronym PECOTA stands for Player Empirical Comparison and Optimization Test Algorithm, and it serves as a sabermetric system designed for predicting the performance of Major League Baseball players. PECOTA goes beyond traditional forecasts, extending its predictions to cover a player's performance in all major categories relevant to typical fantasy baseball games. Additionally, the system provides forecasts for production in advanced sabermetric categories developed by Baseball Prospectus, such as VORP and EqA. PECOTA also includes forecasts for several summary diagnostics, including breakout rates, improvement rates, and attrition rates. Moreover, the system offers insights into the market values of players, rounding out its comprehensive approach to player performance forecasting.

- FanGraphs.com

FanGraphs.com stands as an extensive online resource offering statistics for every player in Major League Baseball history. Navigating the FanGraphs homepage reveals a wealth of content, including articles, statistical reports, coverage of baseball history, and insights into current issues and events such as games, series, injuries, forecasts, player profiles, baseball finance, and the player marketplace. At the core of FanGraphs is the Library, curated by Neil Weinberg, serving as an encyclopedia of Sabermetric statistics and principles. This section provides a deep dive into the analytical aspects of baseball, enriching the user experience with comprehensive insights into the statistical foundation of the game.

- KBO official website

The official website of the Korea Baseball Organization (KBO). The website is a go-to source for KBO League info, offering match schedules, detailed records, player updates, and media content. It highlights legendary players, provides club history, and includes important notices. The "About KBO" section covers organization details, major events, and safety policies. For Futures League enthusiasts, there's info on schedules and rankings. It's a comprehensive hub for all things KBO and baseball.

- MLB.com

MLB.com, the official website of Major League Baseball, is managed by Major League Baseball Advanced Media, L.P., a subsidiary of MLB. The site serves as a comprehensive hub for baseball enthusiasts, offering a wealth

of information such as news, statistics, and sports columns. Functioning as both an informative and commercial platform, MLB.com provides online streaming services, allowing paying subscribers to access video and audio broadcasts of all Major League Baseball games. Additionally, the site offers a free "gameday" feature, delivering near-live streaming box scores for baseball games. Beyond digital content, MLB.com serves as a marketplace for official baseball merchandise and facilitates the purchase of tickets to baseball games. The site also hosts fantasy baseball leagues and auctions for baseball memorabilia. Collaborating with HB Studios, MLB.com has been involved in the development of recent R.B.I. Baseball installments.

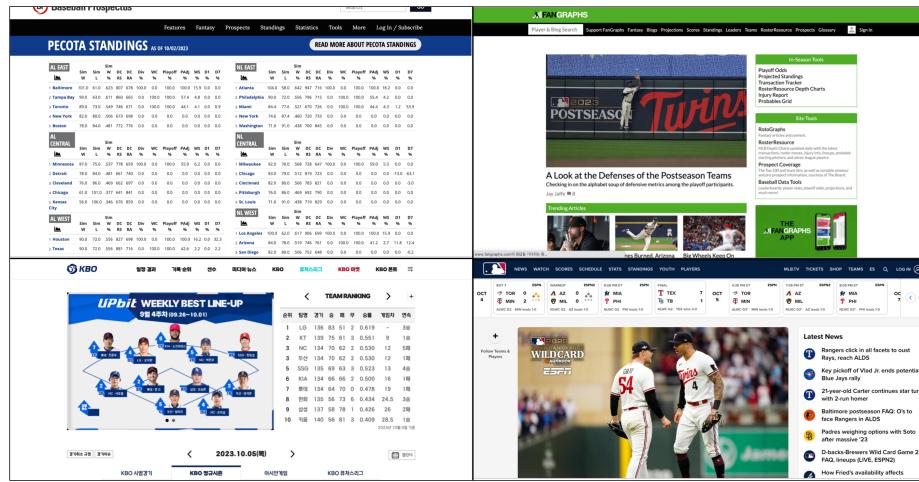


Fig. 1. Existing service

## 4 Problem Statement and Proposed Solution

### 4.1 Problem and Targets

**Problems** First, people who find advanced analytical techniques are confused. The indicators such as  $wRC^+$  and  $BABIP$  are not as intuitive and familiar as the existing simple ones.

$$BABIP = \frac{H - HR}{AB - K - HR + SF}$$

$$wRC^+ = \frac{(wRAA/PA + LgR/PA) + (LgR/PA - ParkFactor \cdot LgR)}{(wRC/PA)}$$

They need additional explanation. They also have difficulty gathering information to analyze players' performance. After finding each player's performance or performance in a specific situation, they must merge themselves. In addition, it takes a lot of effort to analyze the information collected. There is no tool anywhere that compares multiple players or compares them with the average performance of a club or a specific group. This moves baseball fans away from baseball data. It acts as a shield for new and casual fans.

**Targets** To solve these problems, setting targets, needs and devised solutions are necessary. We have set the following three conditions for the target. First, targets are dedicated and highly engaged fans. Fig.2. Second, targets are fans who are interested in the baseball analysis. Fig.3. Third, targets are fans who use online media. Fig.4. The graph below each description show that each target is a high percentage of all baseball fans. This service's core targets are deeply interested in baseball and want easy access to a variety of game and player's data. There can be more targets. One is fans including newbies who want to enjoy baseball and baseball analysis easily. The other is baseball sports industry officials who seek for new fans to expand the market scale.

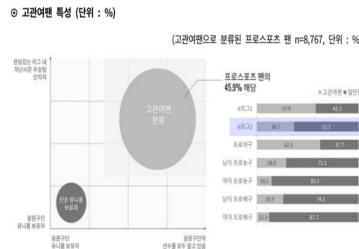
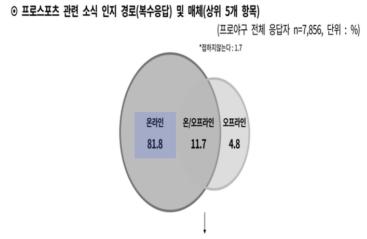


Fig. 2. Highly engaged fan[1]



Fig. 3. Fans who are interested in analysis[1]



**Fig. 4.** Fans who use online media[1]

## 4.2 Solutions and Effects

**Solutions** To address these challenges, we have decided to create web platform BaseballMetrics. This Web Platform offers a range of user-friendly indicators that are related to the baseball and simulation game with our new analytical model.

First, this service provides comprehensible explanation of the existing baseball related indicators. And these indicators are also related to sabermetrics. There are many difficult terms. WHIP, BABIP, wRC+, etc. In baseball broadcasts and news, these terms are often used, but most viewers and readers do not understand their meanings. Furthermore, these complex terms become obstacles for new fans to develop an interest in baseball. So, this website gives information about these terms with wiki style and with easy explanations. Of course, there are existing wiki websites (e.g., Namuwiki, Wikipedia) that cover these topics. However, we aim to differentiate ourselves by mentioning exceptional players associated with these metrics and simplifying the calculation methods for easy comprehension.

Second, this service provides a simulation game with our new analytic model. With sabermetrics, it is thought that can make new algorithm and prediction model. With this model, new simulation game can be made. The user can select 9 hitters. These hitters will be listed up with their sabermetrics data. And then game gives score according to our prediction model.

There are two reasons why this platform serves this game service. Firstly, in Baseball game, there are many breaktime, middle and end of innings, pitcher change, injury, etc. People who enjoy baseball with internet, these times are very boring. So, with this web service these breaktimes will pass by smoothly. Furthermore, through this game, users will gain a better understanding of sabermetrics. After the game, algorithm will give information of our prediction, and recommended team. So they can do this game one more time after analyzing their line-up.

Lastly, this service calculates and provide magic number. In sports, the "magic number" is the specific number of wins a team needs to secure a playoff spot or win a championship, with lower numbers indicating they are closer to achieving that goal. Most service or news provides magic number of 1st team. However, many fans are eager to know our team's magic number to reach the postseason. Or some freak fans want to know our team's magic number to reach the bottom (This is because the last-place team gets the first overall pick in next year's draft.). So this service will calculate and serve the advanced magic number to users.

**Effects** With this solution, two effects can be expected.

First of all, this platform can give more fun for current fans. Simulations make baseball more enjoyable for existing fans by helping them understand and analyze the coach's choices. This increases their satisfaction and involvement with the game. And also, they can spend breaktime smoothly. The break time in the middle of the baseball game won't be a boring time anymore.

Secondly, new fans can feel fun to baseball, and they can easily access to baseball. Learning about different baseball statistics and indicators makes watching games more exciting for new fans. It also attracts more people to become interested in the sports.

This initiative will exert a favorable influence on the sports industry, by reducing the obstacles to accessing baseball-related information for a broader audience.

**Limitations** However, there can be some limitations.

1. Algorithm Complexity: Challenges arise when applying algorithms to various sabermetrics due to their intricate nature.
2. Algorithm Validation: Ensuring the accuracy and reliability of results derived from internally developed algorithms presents difficulties.

These two limitations can be minimalized by thoroughly validating them through historical baseball data up to the present.

3. Professional-Level Data: At the professional level, factors like ball rotation speed and movement are critical, but gathering precise records for such parameters proves challenging.

It is difficult to address this limitation at the moment. There are two reasons. First, these data are difficult to collect and have very high value, clubs do not disclose them very well. Secondly, there has not been much research on evaluating players through these data. If this service is published and people are interested in it, it can be developed into a service that uses this high-level data in the future.

## 5 Planning in Detail

### 5.1 Roles

The roles were largely divided into UX/UI design, front-end, back-end development, and algorithm development. To facilitate smooth communication between frontend and backend, an additional role was added that is responsible for both frontend and backend.

Name	Chaewon Go	Hyeonchang Shin	Changwoo Shim
UX/UI	O		
FrontEnd	O		O
BackEnd		O	O
Algorithms	O	O	O

Table 1. Roles

### 5.2 Schedule

Considering the estimated time required for each part, we planned the schedule as shown in the figure below.

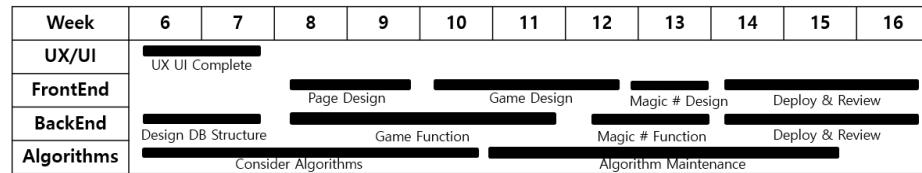


Fig. 5. Schedule

## References

1. 2022 프로스포츠 관람객 성향조사 보고서\_프로야구 보고서 — KBO 팬 7856명(한국프로스포츠협회)
2. Albert, Jim (2010). "Sabermetrics: The Past, the Present, and the Future"
3. Ball, Andrew (January 17, 2014). "How has sabermetrics changed baseball?". Beyond the Box Score.
4. MLB Glossary - Weighted Runs Created Plus (wRC+)
5. MLB Glossary - Batting Average on Balls in Play (BABIP)
6. MLB Glossary - Walks And Hits Per Inning Pitched (WHIP)
7. Barnes, Nate. "Ultimate Guide to Value Over Replacement Player - VoRP". Sporting Charts.

8. PECOTA Standings
9. FanGraphs Baseball — Baseball Statistics and Analysis
10. KBO 홈페이지
11. MLB.com — The Official Site of Major League Baseball